Simultaneous Dependence of the Earthquake-Size Distribution on Faulting Style and Depth

A. Petruccelli[1,2], P. Gasperini[2], T. Tormann[1,3], D. Schorlemmer[4], A.P. Rinaldi[1], G. Vannucci[5], and S. Wiemer[1]

[1] Swiss Seismological Service, ETH Zurich, Zurich, Switzerland, [2] Dipartimento di Fisica e Astronomia, University of Bologna, Bologna, Italy, [3] Now at PartnerRe, Zurich, Switzerland, [4] Helmholtz Centre Potsdam, GFZ Potsdam, Potsdam, Germany, [5] Istituto Nazionale di Geofisica e Vulcanologia, Bologna, Italy

Correspondence to: A. Petruccelli, antonio.petruccelli@sed.ethz.ch

Abstract

We analyze two high-quality Southern Californian earthquake catalogues, one with focal mechanisms, to statistically model and test for dependencies of the earthquake-size distribution, the *b* values, on both faulting style and depth. In our null hypothesis, *b* is assumed constant. We then develop and calibrate one model based only on faulting style, another based only on depth dependence and two models that assume a simultaneous dependence on both parameters. We develop a new maximum-likelihood estimator corrected for the degrees of freedom to assess models' performances. Our results show that all models significantly reject the null hypothesis. The best performing is the one that simultaneously takes account of depth and faulting style. Our results suggest that differential stress variations in the Earth's crust systematically influence *b* values and that this variability should be considered for contemporary seismic hazard studies.

1 Introduction

The frequency-magnitude distribution of earthquakes obeys a negative exponential model with increasing magnitude *M*, often referred to as the Gutenberg-Richter (GR) law (Gutenberg & Richter, 1944), with the slope *b* of the decay (*b* value) ranging typically between 0.5 and 1.5 (Wiemer & Wyss, 2000). This empirical relationship is one of the cornerstones of seismic hazard assessments, since it enables more frequent, smaller events ($b > 1$) to be related to infrequent, damaging events ($b < 1$). The *b* value has been empirically shown to be inversely related to differential stress $\Delta\sigma$. This finding has been confirmed in the laboratory (Amitrano, 2003; Goebel et al., 2013; Kwiatek et al., 2014) but is also consistent with observations of natural and induced earthquakes (Bachmann et al., 2012; Gulia et al., 2018; Gulia & Wiemer, 2010; Schorlemmer et al., 2005; Tormann et al., 2015).

There are two major gradients of differential stress in the Earth's crust, and both have been shown to also correlate with changes in the *b* value:

1. Differential stress *increases with depth* in the Earth's brittle crust, until it reaches more ductile regimes (Brace & Kohlstedt, 1980; Kirby, 1980). Mori and Abercrombie (1997), Gerstenberger et al. (2001), and most

recently Spada et al. (2013) have shown that the b value decreases with increasing depth, as expected, then increases again as one approaches the brittle ductile transition zone. Scholz (2015) has related this observation to the average stress gradient in the crust.

2. Differential stress *changes systematically between faulting regimes*. The classical theory of faulting (Anderson, 1905) assumes that various states of stress are associated with different tectonic regimes around the source volume, with higher stress for compressive environments and lower stress for extensional environments. Schorlemmer et al. (2005) firstly evidenced differences in b for different tectonic regimes on rake angle λ of the focal mechanisms (FMs). Recently, a functional trend of the b value with λ was statistically tested on a global scale (Petruccelli et al., 2018) and described by the harmonic trend $b \approx -\sin \lambda$, whereby the minimum value of b is for thrust faulting regimes ($b \approx 90°$), the maximum value for normal faulting regimes ($b \approx -90°$), and intermediate values for strike-slip (SS) faults ($b \approx 0, \pm 180°$).

Despite the fact that b values in the lab are known to depend on stress, seismotectonic and hazard analyses are not yet using these important physical constraints as input. We believe that this is so because there is a lack of calibrated and validated models and a lack of trust in the robustness. Our paper addressed the gap and is the first to systematically evaluate and test the impact of each of these gradients, as well as the combination of both, on the b value. We first extend the method for likelihood estimation and then define and analyze four models, or hypotheses, of different b value dependencies, with different degrees of complexity (i.e., number of free parameters n for each model). We then test the performance of each model, adjusted for degrees of freedom, and interpret the results. Finally, we comment on implications for hazard assessment.

2 Data and Methodology

2.1 Southern California Data Sets

FM catalogue: For our study, we used one of the largest, highest-quality FM data sets available worldwide, namely, the one from Southern California (SC; Hauksson et al., 2012; Yang et al., 2012), including data from 1981 to the end of September 2016. SC FM quality is divided into four classes, ranging from A (*best*) to D (*worst*), according to the azimuthal gap and the mean nodal plane uncertainty. For our analysis, our choice to retain only the first two classes, A and B, resulted in a reliable analysis (Figure 1). For earthquakes occurring at the edge of the seismic network (Classes C and D), the quality factor can be systematically biased, resulting in a lower probability (less than 50%) of detecting well-constrained mechanisms (Yang et al., 2012; see Figure 6).

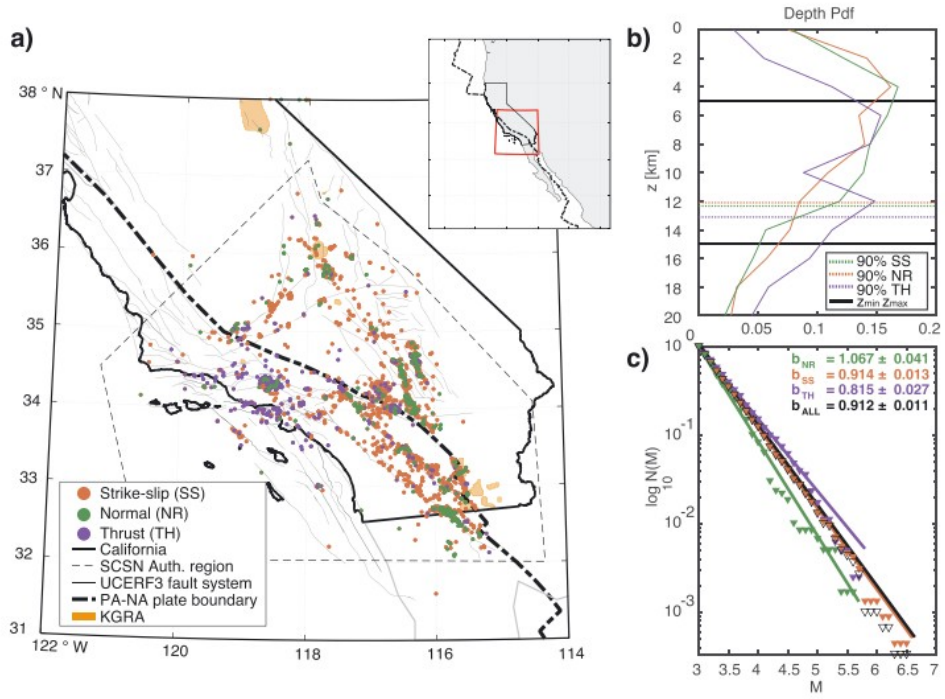

Figure 1. SC FM data set (1981–2016, z = 5–15 km, Mc = 3, A and B mechanisms). (a) SC FM map. Almost all data fall inside the authoritative region of Southern California Seismic Network (SCSN—see the dashed line). Orange shading indicates the Known Geothermal Resource Area (KGRA) of SC. UCERF3: Uniform California Earthquake Rupture Forecast Version 3. NA: North America. PA: Pacific. (b) Hypocentral depth probability density functions (PDFs; depth binning at 2 km). The 90% of seismicity for each family denotes the seismogenic depth border. z_{min} z_{max}: minimum and maximum selected depth for further analyses. c) Normalized frequency-magnitude distributions of data in (a). White, black-outlined empty markers represent the sum of all three families.

In line with Yang et al. (2012), we considered normal faulting mechanisms (NR) to be events with a rake angle of $-135° \leq \lambda \leq -45°$ (green dots in Figure 1a), thrust mechanisms (TH) to be events with a rake angle of $45° \leq \lambda \leq 135°$ (blue dots), and classified all other events as SS faults (red dots). Since most of planes of SS events are right lateral ($\lambda \sim \pm 180°$), we computed the alternative plane for each earthquake and then stacked the rake angles to obtain a complete set of possible rakes λ. The rake frequencies stacking procedure (for which each earthquake was counted twice) for b value analysis—already applied on a global scale (Petruccelli et al., 2018)—lowered uncertainty, since it was always consistent with b analysis for principal/alternative planes. We considered the crustal depth range at which SC seismicity most frequently occurs (z = 5–15 km, Figure 1b), excluding the shallowest earthquakes, as proposed by Spada et al. (2013), and also the deeper events, where a linear rheology, and thereby linearity between b and differential stress, might no longer be valid. The overall data set was complete for M above 2.5 (Yang et al., 2012). Since we needed to divide our original data set according to tectonic styles, we used a higher completeness threshold, confirming each style-data set to be complete over M = 3 (see frequency-magnitude distribution in Figure 1c). We tested our results for the sensitivity of this choice and also analyzed the influence of the quality criteria and missing events from faulting style (see supporting information Text S1).

To verify the consistency of the *b* value with depth gradient between catalogue subsets with FM and the full catalogue without FM, we also analyzed the SC data set of Hauksson et al. (2012; HS) from 1981–2016 for the same study region and cut at a conservative completeness of $Mc = 2.5$ (Hauksson et al., 2012; Schorlemmer & Woessner, 2008). Accordingly, we did not need to consider a potential bias from changing ML definitions that affect lower-magnitude events during this time period (Tormann et al., 2010). No quality criteria were applied, and for this catalogue we used the full depth range, that is, 0–20 km.

2.2 Likelihood Estimators for Model Evaluation

In statistical seismology, the maximum likelihood estimation (MLE) method is the standard approach used to compute the GR *b* value. The *b* values are computed by maximizing the logarithm of the likelihood function $\mathscr{L} = \prod_{i=1}^{N} f(M_i)$ (where *f(M)* is the probability density function) that is the log likelihood function $L = \ln\mathscr{L}$ for the GR model (see supporting information Text S2). The result of this analytical computation leads to the widely used analytical formula for estimating *b* values (Aki, 1965) using the binning correction set out in of Bender (1983):

$$b = \frac{\log_{10} e}{\overline{M} - \left(M_C - \frac{\Delta M}{2}\right)} \quad (1)$$

where $\overline{M}$ is the mean magnitude, M_C is the minimum magnitude of completeness, and ΔM is magnitude binning. If *b* is expressed as a function of physical variables (see Text S2), the analytical formula 1 becomes unsuitable, since the number of parameters to estimate inevitably increases. To overcome this problem, we developed a novel modified MLE application to numerically estimate the "best fitting" point $\widehat{b} = \max[L]$ from a general GR log likelihood function (see Texts S1 and S2):

$$L(b) = \ln\left\{\prod_{i=1}^{N} f M_i|b\right\} = \sum_{i=1}^{N} \ln\left\{b \ln 10\ 10^{-b\left[M_i - M_c - \frac{\Delta M}{2}\right]}\right\} \quad (2)$$

where *b* is now a function with *n* free parameters p_j. It is clear that if $n = 1$, the maximum of such a function can be obtained both analytically and numerically, and the two estimations coincide. An increase in the number of free parameters p_j ($j = 1, 2 \ldots, n$) results in an increase in the complexity of the modeled $b(p_1, p_2, \ldots p_n)$ used to explain the physical reality. So in equation 2 we could explicitly express varying *b* value dependence (on depth and tectonic styles separately or simultaneously) in equations $b(p_1, p_2, \ldots p_n)$. Next, we could maximize the log likelihood and obtain a score for each model. Finally, we could evaluate how well each model fit (see next paragraph) and rank them according to some statistical criteria (Akaike, 1974; Cavanaugh, 1997; Neyman & Pearson, 1933). Standard errors on the optimal parameters of each model were computed as the square roots of the diagonal elements in the variance-covariance matrix.

2.3 Goodness-of-Fit and Significance in Different Models

In general, the higher the number of model parameters, the better the model fit the data, hence the larger the likelihood (see Table 1). However, to select the best out of an available set of possible models, we had to account for the number of free parameters for each model by considering an appropriate penalty term for more complex models as opposed to simpler ones. This was achieved by applying the second-order Akaike information criterion (AIC), often referred to as "corrected" AIC_c (Akaike, 1974; Cavanaugh, 1997):

$$AIC_c = [-2\log L + 2n] + \frac{2n(n+1)}{N-n-1} = AIC + \frac{2n(n+1)}{N-n-1} \quad (3)$$

which, in addition to the log likelihood score (positive), also takes account of sample size N by increasing a relative penalty term for complex models (with a high number of free parameters n) for small data sets (low N). The AIC_c scores were such that the best model had the lowest AIC_c (see Table 1).

Table 1
Summary of Statistical Analyses on the b value Models for FM Data Set

									L_R test (p value, %)			
									⇒ Models M_A ⇒			
b value	n		Model equation	Estimated parameters	$L(\hat{b})$	AIC_c	w	R	0	1	2	3.1
⇓ Models M_B ⇓												
Constant	1	0	$b = b_0$	$b_0 = 0.912 \pm 0.012$	−1,539.7	3,081.4	0%	Fifth	—	—	—	—
Depth	2	1	$b(z) = b_0 - \frac{db}{dz}z$	$b_0 = 1.067 \pm 0.012$ $\frac{db}{dz} = 0.017 \pm 0.001$ km^{-1}	−1,532.9	3,069.8	<0.1%	Fourth	0.02	—	—	—
Faulting style	3	2	$b_{STYLE} = b_{STYLE}$	$b_{NR} = 1.067 \pm 0.043$ $b_{SS} = 0.914 \pm 0.013$ $b_{TH} = 0.815 \pm 0.028$	−1,527.4	3,060.7	5.0%	Third	<0.01	0.09	—	—
Faulting style and depth	3	3.1	$b(\lambda, z) = b_0 - \kappa\frac{2\mu\rho g}{\sqrt{\mu^2+1}-\mu\sin\lambda}z$	$b_0 = 1.046 \pm 0.012$ $\kappa = (8.58 \pm 0.63)*10^{-4}$ $\mu = 0.524 \pm 0.036$	−1,525.4	3,056.9	32.5%	Second	<0.01	0.09	—	—
	6	3.0	$b_{STYLE}(z) = b_{0,STYLE} - \left(\frac{db}{dz}\right)_{STYLE}z$	$b_{0,NR} = 1.101 \pm 0.043$ $\frac{db}{dz_{0,NR}} = 0.004 \pm 0.005$ km^{-1} $b_{0,SS} = 1.046 \pm 0.013$ $\frac{db}{dz_{0,SS}} = 0.015 \pm 0.001$ km^{-1} $b_{0,TH} = 1.023 \pm 0.028$ $\frac{db}{dz_{0,TH}} = 0.021 \pm 0.003$ km^{-1}	−1521.8	3,055.6	62.5%	First	<0.01	0.02	1.07	6.58

Note. From top to bottom (rows): Models M_B (see section 2.3, an alternative hypothesis being that the increase in n is statistically significant). From left to right (columns): b value hypothesis, number of free parameters n, models numbering, models equations, best fit parameters, log likelihood values, AIC_C values and corresponding weighting w and ranking R, L_R ratio test for p values (p = probability of being wrong if the null hypothesis M_A = best is rejected at 5% significance) between M_B models and M_A models (the null hypothesis being that the increase in n is not statistically significant). Note how L_R test were only performed on models of different complexity, such that n (M_B) > n (M_A).

Generally, two candidate models with different degrees of complexity (e.g., M_A and M_B, where A has less free parameters than B) could be evaluated based on the performance of a null hypothesis H (M_A) with respect to the alternative H (M_B). This concept can be expressed using the log likelihood ratio $L_R = -2\ln\frac{\mathcal{L}M_A}{\mathcal{L}M_B} = 2[LM_B - LM_A]$ (Neyman & Pearson, 1933). If the L_R criterion test was passed, the null hypothesis, under which the "simple" M_A performed better than the "complex" M_B, if L (M_B) was sufficiently larger than L (M_A) ($L_R < k$, where k is a certain significance threshold) could be rejected. To determine whether such a difference in likelihood scores was statistically significant, assuming that the L_R statistic approximately followed a χ^2 distribution (Wilks, 1938), we could then compute the critical value (p value) of the test statistic from standard statistical tables making the degrees of

freedom equal to the difference in the number of free parameters for the two candidate models (see Table 1, right-hand columns).

Furthermore, even AIC differences could be turned into log likelihood ratios according to (Burnham & Anderson, 2002) by computing the weight of evidence in favor of *i*th model as the actual best model for the situation at hand:

$$w_i = \frac{L_R(M_i)}{\sum_{j=1}^{N} L_R(M_j)} = \frac{\exp\left[-\frac{1}{2}AICc_i\right]}{\sum_{j=1}^{N} \exp\left[-\frac{1}{2}AICc_j\right]} \quad (4)$$

3 Models and Results

We develop, calibrate, and subsequently evaluated three categories of models: a no-variable model, single-variable models (Figures 1 and 2) and multivariable models (Figure 3). The no-variable was Number 0, the single-variable models were Numbers 1 and 2 (see Table 1 and Figure 2), and the multivariable models were Numbers 3.0 and 3.1 (see Table 1, Figure 3, and supporting information Text S3). We then analyzed the performance of each model and ranked them accordingly. For single-variable models, we also reanalyzed the *b* value using depth and faulting style dependencies established in the literature for both our data sets against the null hypothesis of a single, constant *b* value.

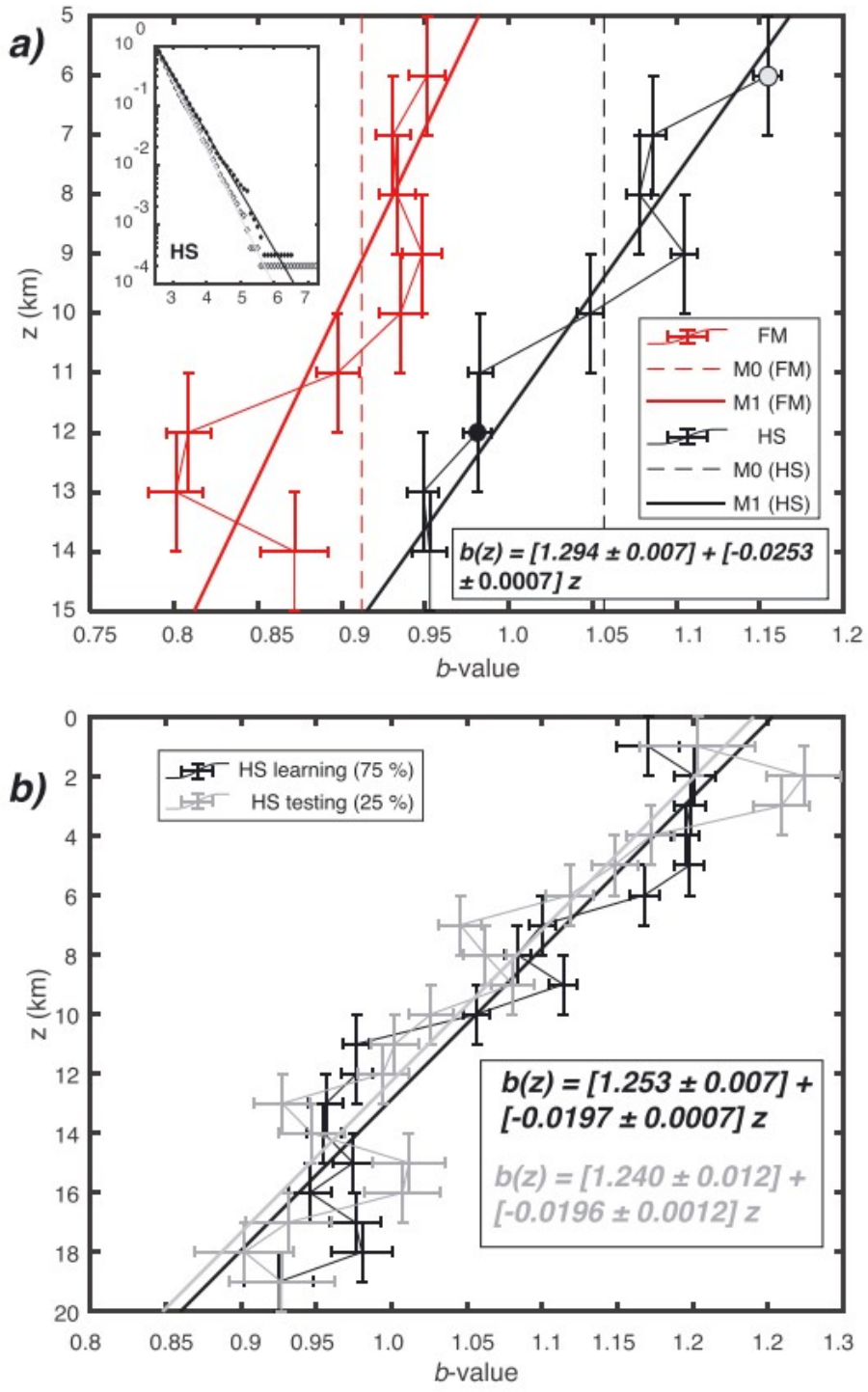

Figure 2. Models 0, 1 maximum likelihood estimation analysis (a) for FM and HS catalogues separately and (b) for HS training (75% data) and testing (25%) subsets only. While M0 offset (dashed lines) and M1 gradients (bold lines) were estimated using all the available data (parameters for FM presented in Table 1, while parameters for HS are presented in the bottom, right-hand corner), the *b* value versus depth error bars (solid lines) were obtained by sampling data inside moving windows (step = 1 km) of constant (2 km) width (vertical error bar) applying the classical maximum likelihood estimation method (Aki, 1965; see supporting information Text S2) with uncertainties regarding *b* (horizontal error bars) given by Shi and Bolt (1982), following Spada et al. (2013). Inset box: HS normalized cumulative frequency-magnitude distributions for indicated *b* values (circles at $z = 6$ and 12 km). FM = focal mechanism.

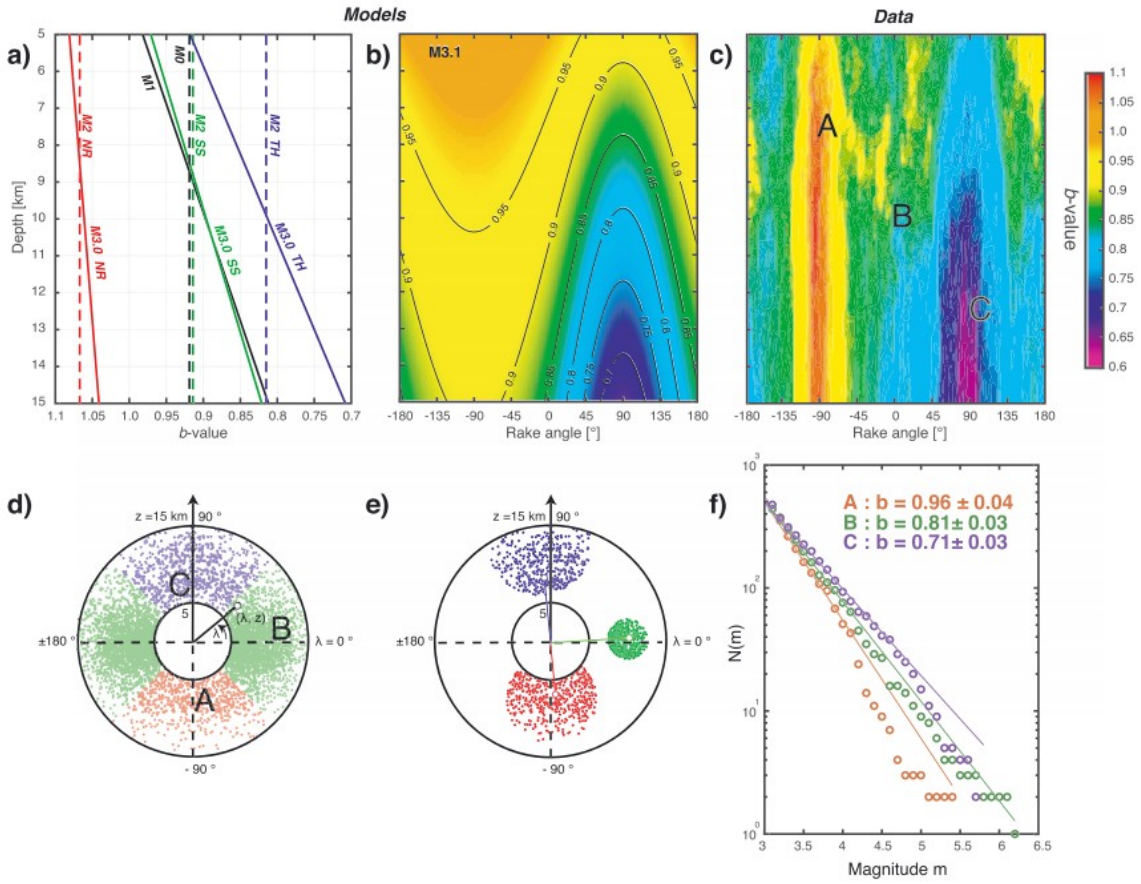

Figure 3. Representation of Models M0–M3 in rake-depth space. (a) The *b* models in depth space. Model 0: black dashed line. Model 1: black solid line. Model 2: colored dashed lines. Model 3.0: colored (same as Model 2) solid lines. (b) Model 3.1: (λ, z) contour (with *b* isolines indicated) and surface (shaded) plots. (c) The *b* value mapping surface into the rake-depth space. Each *b* value was computed (using equation (1)) by sampling its 500 neighbors (see d and e). A, B, and C indicate sampled distribution point (see d and e). (d) Polar parametrization (λ, z) for nearest neighbor algorithm and Points A, B, C in (c). The parametrization entailed setting $b(\lambda, z)_H = z \cos\lambda$ and $b(\lambda, z)_V = z \cos\lambda$ (H = horizontal component, V = vertical component). Parameter z is the depth vector (acting as a radius), while rake indicates the orientation, and then focal mechanism. The doughnut-shaped arises from the exclusion of shallower events. (e) Events sampled in depth-rake space around Points A, B and C, shown in (d). (f) Frequency-magnitude distributions for the Samples A, B, and C are shown in (c)–(e).

Model 0 (M0). This model assumed a constant *b* value for the entire data set and served as the null hypothesis. M0 estimates only one parameter and can be represented in Figures 2a and 3a as a straight line. The numerically estimated value for FM, $b = 0.91$, coincides with the analytical value obtained using the method presented by (Aki, 1965; Figure 1c), with similar errors, illustrating that the two approaches are interchangeable. The lowest log likelihood, highest AIC_c and a very low weight make this the worst performing model.

3.1 Single-Variable Models

Model 1 (M1). This model assumed that *b* values depend linearly on depth (or differential stress), in line with Scholz (2015) and Spada et al. (2013), and can be illustrated as a line with a negative gradient in Figures 2a and 3a. M1 has two free parameters to estimate, the intercept and the slope with respect to depth. On FM data, compared to M0, the log likelihood increases, AIC_c decreases, and the L_R *p* value is significant. Accordingly, M1 significantly

improves the description of the data set. The depth slope is similar to the one that would arise from the dependence on differential stress hypothesis by Scholz (2015; see supporting information Text S4 and Table S1).

M1 was also evaluated for the full HS catalogue (Figure 2a) with comparable results: Its M1 log likelihood (−2,327.1) is better than M0 (−2,369.6), M1 AIC_c (4,658.2) is lower than the M0 (4,741.3), and the p value is very small, which makes a highly significant difference. Because the HS data were very rich, we also evaluated the temporal stability of the results (Figure 2b) by subdividing the data set into a first period (75% of the data) and second period (the remaining 25%), and ended up with almost identical equations for both data sets (Figure 2b). The equations also remained consistent when changing the partition sizes of the training and testing data sets (even at 50 ÷ 50). The overall depth gradient of b does not depend on temporal selections. Applying a pseudoprospective approach (i.e., using the learning period to make forecasts, then evaluating the likelihood for the testing period), we found that L(M0) = −552.98, L(M1) = −547.15, giving Model 1 a relative weight of 99.2%.

Model 2 (M2). This model assumed that the b value is constant with respect to depth but varies for different faulting styles. M2 has three free parameters to estimate: the b values for the three styles. The three faulting styles' b values are represented in Figure 3a as straight lines. The values we obtained were consistent with the estimations provided by Yang et al. (2012): The AIC_c was lower than for M0 and M1, and the difference carried a significant p value, so faulting style proved to be a more relevant parameter than depth determining b values.

3.2 Multivariable Models

Multivariable models assume that b depends on both faulting style and depth (Figures 3a and 3b). As a first step, we investigate for the first time simultaneously the depth and faulting style dependency of the b value in the FM catalogue. In a Euclidean rake-depth (λ, z) space (Figure 3c), we used polar parametrization (see Figure 3d) to find the closest 500 events (Figure 3e) and computed the corresponding b value (Figure 3f). The results are presented in Figure 3c, clearly showing both rake and depth dependence. Building on this finding, we develop the first multivariable models to describe these dependencies.

Model 3.0 (M3.0). Each tectonic style can be described by taking a different equation of decreasing b value as a function of depth (or differential stress) for a total of six free parameters. This is represented in Figure 3a as three lines with different slopes. This model generates the highest L value overall, the lowest (and hence best) *AICc*, a significant L_R p value with respect to all other models and a relative weight in excess of 60%, making it the best predictive model. In line with faulting theory (Anderson, 1905), a steeper depth gradient was found for TH, an intermediate depth gradient for SS, and a lower depth gradient for NR. According to our equations (see supporting

information Text S5), the *b* value decreases by about 0.04 every 10 km for NR, by about 0.15 every 10 km for SS, and by about 0.2 every 10 km for TH.

Model 3.1 (M3.1). This model assumed a sinusoidal dependence of *b* on rake, similar to that proposed by Petruccelli et al. (2018) but we now add the depth dependency with *b* stress gradient κ, modulated by sin λ:

$$b(\lambda, z) = b_0 - \kappa \frac{2\mu\rho g}{\sqrt{\mu^2 + 1} - \mu \sin\lambda} z \quad (5)$$

Equation 5 offers a generalized view of the simultaneous dependencies of *b* value on tectonic styles (i.e., rake angle λ) and depth and is plotted in Figure 3b. M3.1 has three free parameters to estimate (offset b_0, overall stress gradient κ, and friction μ). The estimated value for the "overall" stress gradient κ is in line with the SS stress gradient k_s of M3.0 (see supporting information Text S4). The estimated best friction value $\hat{\mu}$ of Model 3.1 was somewhat lower than the assumed value usually assumed (0.6–1.0) but consistent with previous estimations (Carpenter et al., 2011; Zoback et al., 1987). However, recent advances infer that the fault zones may actually present much lower frictional coefficients (Collettini et al., 2019). M3.1 fit the data only slightly worse than M3.0, making it the second-best model. However, in line with Burnham and Anderson (2002), the low likelihood ratio between M3.0 and M3.1 (1.92) indicated a weak support for Model 3.0 being best. In other words, uncertainty in choosing between Models 3.0 and 3.1 is likely to be high. Conversely the high-likelihood ratios between M3.0 and M3.1 and the other models (larger than 12 and 6, respectively) permit the conclusion that other models (0, 1, and 2) are very unlikely to perform best.

Since Models 3.0 and 3.1 are not independent, we can also compute the relative weights between Models M0, M1, M2 and only M3.0 (around 0.0%, 0.1%, 7.0%, and 93.0%) as well as only M3.1 (0.0%, 0.1%, 12.0%, and 88.0%). These could be used as weightings in a probabilistic seismic hazard analysis (PSHA) logic tree where models are assumed to express epistemic uncertainties and should be ideally mutually exclusive and cumulatively exhaustive.

4 Discussion and Conclusions

What is the best predictor of the *b* value of the frequency-magnitude distribution in SC (Figure 1): depth, faulting style, or a combination of both these parameters? To address this important yet until today unanswered question, we first needed to create a generalized MLE enabling model parameters to be estimated, then develop and calibrate adequate models, and finally compare the models' performance. We defined and compared four different models (Figures 2 and 3 and Table 1), or—with varying degrees of freedom—hypotheses that expressed different dependencies. We also compared them to the null hypothesis of a constant *b* value.

All models rejected the null hypothesis (M0), so it can be assumed that *b* = *const*. is a comparatively poor model and should not be considered in hazard

studies. A very clear and statistically highly significant depth dependence for *b* values (M1) emerged from both the FM and HS data and also persisted over time. More hazard studies should take account of this *b* depth gradient, which has previously only rarely been considered. An even better predictor of *b* values was faulting style (M2), which significantly outperformed depth dependence. For hazard-related studies, faulting style was sometimes linked to different ground-motion prediction equations but never explicitly to *b* values. However, our findings cannot be immediately applied in hazard assessment because we assessed faulting style dependence for regional sets of earthquakes, not individual source zones. It is worth noting that depth and faulting style are usually correlated. The Anderson's (1905) theory of faulting relates the faulting style to the differential stress and consequently to depth. Indeed, globally but also for California, thrust faulting requires large differential stress and it occurs usually at greater depth than normal faulting. Another limitation of our work is that the subset of events with high-quality FM we analyzed is relatively small, and FM quality itself may depend systematically on magnitude, faulting style and depth, potentially biasing our analysis. Our quality check (see Text S1) and the overall consistency of the results with other data sets, other regions and a physical theory suggested no such bias, though.

The best predictor by far was found to be a combination of both faulting style and depth (M3.0 and M3.1). In Figure 3c, a first ever depiction of dependence on depth and rake neatly illustrates that while rake is a dominating factor, depth also plays a role. The difference between M3.0 and 3.1 is quite small so either one is a feasible approach for modeling depth and rake dependence. To the best of our knowledge, this study was the first to simultaneously model and examine both depth and faulting style dependence. It also provided yet further confirmation that *b* values' laboratory-observed stress dependency also applies in the Earth's crust. We used the so-called Akaike weights to quantify the relative importance (equation 4) but also to suggest how our model results could be employed in a logic tree-approach common in PSHA to express epistemic uncertainties (Woessner et al., 2015). While M0 received zero weight, both Models 3 scored weightings of 95%, suggesting that they presented data in the best way by far. There are different ways to include our results in future PSHA studies. For example, the a priori, overall *b* value often assigned to a region, or a so-called super zone (Woessner et al., 2015) can be dependent on the preferred faulting style, or on depth. Contemporary hazard studies will compare the a priori value to local (area source or zoneless) estimates and update the prior using the uncertainty in the local estimate in a Bayesian sense (e.g., Broccardo et al., 2017; Wiemer et al., 2009; Woessner et al., 2015). Likewise, hazard calculations commonly use a depth-dependent activity rate, and this could be readily extended to also use a depth-dependent *b* value.

We did not analyze spatial or temporal variations in *b* values, which are often subject to steep gradients (Tormann et al., 2012, 2015) that can be predictive (Schorlemmer, Wiemer, & Wyss, 2004, Schorlemmer, Wiemer, Wyss & Jackson, 2004; Hiemer & Kamer, 2016; Gulia et al., 2018). Our goal was to investigate the most basic gradients but in a statistically rigorous way. Future work should consider even more complex space-time-depth-rake dependencies, ideally in a conceptual framework of stresses in the Earth's crust.

Acknowledgments

For FM catalogue data and for HS data, see the websites, respectively (http://scedc.caltech.edu/research-tools/alt-2011-yang-hauksson-shearer.html and http://scedc.caltech.edu/research-tools/alt-2011-dd-hauksson-yang-shearer.html). We do thank the Editor and two anonymous reviewers for the comments that improved our work.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716– 723. https://doi.org/10.1109/TAC.1974.1100705

Aki, K. (1965). Maximum likelihood estimate of *b* in the formula log*N* = *a* − *bM* and its Confidence Limits. *Earthquake Research Institute, University of Tokyo*. Bull. Earthquake Res. Inst. Univ. Tokyo.

Amitrano, D. (2003). Brittle-ductile transition and associated seismicity: Experimental and numerical studies and relationship with the b value. *Journal of Geophysical Research*, 108(B1), 2044. https://doi.org/10.1029/2001JB000680

Anderson, E. M. (1905). The dynamics of faulting. *Transactions. Edinburgh Geological Society*, 8(3), 387– 402. https://doi.org/10.1144/transed.8.3.387

Bachmann, C. E., Wiemer, S., Goertz-Allmann, B. P., & Woessner, J. (2012). Influence of pore-pressure on the event-size distribution of induced earthquakes. *Geophysical Research Letters*, 39, L09302. https://doi.org/10.1029/2012GL051480

Bender, B. (1983). Maximum likelihood estimation of *b* values for magnitude grouped data. *Bulletin of the Seismological Society of America*, 73(3), 831–851. Retrieved from. http://www.bssaonline.org/cgi/content/abstract/73/3/831

Brace, W. F., & Kohlstedt, D. L. (1980). Limits on lithospheric stress imposed by laboratory experiments. *Journal of Geophysical Research - Solid Earth*, 85(B11), 6248– 6252. https://doi.org/10.1029/JB085iB11p06248

Broccardo, M., Mignan, A., Wiemer, S., Stojadinovic, B., & Giardini, D. (2017). Hierarchical Bayesian modeling of fluid-induced seismicity. *Geophysical*

Research Letters, 44, 11,357– 11,367. https://doi.org/10.1002/2017GL075251

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach*, (2nd ed., Vol. 34, pp. 7389– 7400). New York. New York Springer: Springer-Verlag. https://doi.org/https://doi.org/10.1016/j.biomaterials.2013.06.009

Carpenter, B. M., Marone, C., & Saffer, D. M. (2011). Weakness of the San Andreas Fault revealed by samples from the active fault zone. *Nature Geoscience*, 4(4), 251– 254. https://doi.org/10.1038/ngeo1089

Cavanaugh, J. E. (1997). Unifying the derivations for the Akaike and corrected Akaike information criteria. *Statistics & Probability Letters*, 33(2), 201– 208. https://doi.org/10.1016/S0167-7152(96)00128-9

Collettini, C., Tesei, T., Scuderi, M. M., Carpenter, B. M., & Viti, C. (2019). Beyond Byerlee friction, weak faults and implications for slip behavior. *Earth and Planetary Science Letters*, 519, 245– 263. https://doi.org/10.1016/J.EPSL.2019.05.011

Gerstenberger, M., Wiemer, S., & Giardini, D. (2001). A systematic test of the hypothesis that the b value varies with depth in California. *Geophysical Research Letters*, 28(1), 57– 60. https://doi.org/10.1029/2000GL012026

Goebel, T. H. W., Schorlemmer, D., Becker, T. W., Dresen, G., & Sammis, C. G. (2013). Acoustic emissions document stress changes over many seismic cycles in stick-slip experiments. *Geophysical Research Letters*, 40, 2049– 2054. https://doi.org/10.1002/grl.50507

Gulia, L., Rinaldi, A. P., Tormann, T., Vannucci, G., Enescu, B., & Wiemer, S. (2018). The effect of a mainshock on the size distribution of the aftershocks. *Geophysical Research Letters*, 45(24), 13,277– 13,287. https://doi.org/10.1029/2018GL080619

Gulia, L., & Wiemer, S. (2010). The influence of tectonic regimes on the earthquake size distribution: A case study for Italy. *Geophysical Research Letters*, 37, L10305. https://doi.org/10.1029/2010GL043066

Gutenberg, B., & Richter, C. F. (1944). Frequency of earthquakes in California. *Bulletin of the Seismological Society of America*, 34, 185– 188.

Hauksson, E., Yang, W., & Shearer, P. M. (2012). Waveform relocated earthquake catalog for Southern California (1981 to June 2011). *Bulletin of the Seismological Society of America*, 102(5), 2239– 2244. https://doi.org/10.1785/0120120010

Hiemer, S., & Kamer, Y. (2016). Improved seismicity forecast with spatially varying magnitude distribution. *Seismological Research Letters*, 87(2A), 327– 336. https://doi.org/10.1785/0220150182

Kirby, S. H. (1980). Tectonic stresses in the lithosphere: Constraints provided by the experimental deformation of rocks. *Journal of Geophysical Research*

Solid Earth (1978–2012), 85(B11), 6353– 6363. https://doi.org/10.1029/JB085iB11p06353

Kwiatek, G., Goebel, T. H. W., & Dresen, G. (2014). Seismic moment tensor and *b* value variations over successive seismic cycles in laboratory stick-slip experiments. *Geophysical Research Letters*, 41, 5838– 5846. https://doi.org/10.1002/2014GL060159

Mori, J., & Abercrombie, R. E. (1997). Depth dependence of earthquake frequency-magnitude distributions in California: Implications for rupture initiation. *Journal of Geophysical Research - Solid Earth*, 102(B7), 15,081–15,090. https://doi.org/10.1029/97JB01356

Neyman, J., & Pearson, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 231(694–706), 289– 337. https://doi.org/10.1098/rsta.1933.0009

Petruccelli, A., Vannucci, G., Lolli, B., & Gasperini, P. (2018). Harmonic fluctuation of the slope of the frequency–magnitude distribution (b-value) as a function of the angle of rake. *Bulletin of the Seismological Society of America*, 108(4), 1864– 1876. https://doi.org/10.1785/0120170328

Scholz, C. H. (2015). On the stress dependence of the earthquake *b* value. *Geophysical Research Letters*, 42, 1399– 1402. https://doi.org/10.1002/2014GL062863

Schorlemmer, D., Wiemer, S., & Wyss, M. (2004). Earthquake statistics at Parkfield: 1. Stationarity of b values. *Journal of Geophysical Research, B Solid Earth*, 109(B12), B12307. https://doi.org/10.1029/2004JB003234

Schorlemmer, D., Wiemer, S., Wyss, M., & Jackson, D. D. (2004). Earthquake statistics at Parkfield: 2. Probabilistic forecasting and testing. *Journal of Geophysical Research*, 109(B12), B12308. https://doi.org/10.1029/2004JB003235

Schorlemmer, D., & Woessner, J. (2008). Probability of detecting an earthquake. *Bulletin of the Seismological Society of America*, 98(5), 2103–2117. https://doi.org/10.1785/0120070105

Schorlemmer, D., Wiemer, S., & Wyss, M. (2005). Variations in earthquake-size distribution across different stress regimes. *Nature*, 437(7058), 539–542. https://doi.org/10.1038/nature04094

Shi, Y., & Bolt, B. A. (1982). The standard error of the magnitude-frequency *b* value. *Bulletin of the Seismological Society of America*, 72(5), 1677– 1687. Retrieved from. http://www.bssaonline.org/content/72/5/1677.abstract

Spada, M., Tormann, T., Wiemer, S., & Enescu, B. (2013). Generic dependence of the frequency-size distribution of earthquakes on depth and its relation to the strength profile of the crust. *Geophysical Research Letters*, 40, 709– 714. https://doi.org/10.1029/2012GL054198

Tormann, T., Enescu, B., Woessner, J., & Wiemer, S. (2015). Randomness of megathrust earthquakes implied by rapid stress recovery after the Japan earthquake. *Nature Geoscience*, 8(2), 152– 158. https://doi.org/10.1038/ngeo2343

Tormann, T., Wiemer, S., & Hardebeck, J. L. (2012). Earthquake recurrence models fail when earthquakes fail to reset the stress field. *Geophysical Research Letters*, 39, L18310. https://doi.org/10.1029/2012GL052913

Tormann, T., Wiemer, S., & Hauksson, E. (2010). Changes of reporting rates in the Southern Californian Earthquake Catalog, introduced by a new definition of ML. *Bulletin of the Seismological Society of America*, 100(4), 1733– 1742. https://doi.org/10.1785/0120090124

Wiemer, S., & Wyss, M. (2000). Minimum magnitude of completeness in earthquake catalogs: Examples from Alaska, the Western United States, and Japan. *Bulletin of the Seismological Society of America*, 90(4), 859– 869. https://doi.org/10.1785/0119990114

Wiemer, S., Giardini, D., Fäh, D., Deichmann, N., & Sellami, S. (2009). Probabilistic seismic hazard assessment of Switzerland: Best estimates and uncertainties. *Journal of Seismology*, 13(4), 449– 478. https://doi.org/10.1007/s10950-008-9138-7

Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, 9(1), 60– 62. https://doi.org/10.1214/aoms/1177732360

Woessner, J., Laurentiu, D., Giardini, D., Crowley, H., Cotton, F., Grünthal, G., Valensise, G., Arvidsson, R., Basili, R., Demircioglu, M. B., Hiemer, S., & Consortium, T. S. (2015). The 2013 European seismic hazard model: Key components and results. *Bulletin of Earthquake Engineering*, 13(12), 3553– 3596. https://doi.org/10.1007/s10518-015-9795-1

Yang, W., Hauksson, E., & Shearer, P. M. (2012). Computing a large refined catalog of focal mechanisms for southern California (1981–2010): Temporal stability of the style of faulting. *Bulletin of the Seismological Society of America*, 102(3), 1179– 1194. https://doi.org/10.1785/0120110311

Zoback, M. D., Zoback, M. L., Mount, V. S., Suppe, J., Eaton, J. P., Healy, J. H., Oppenheimer, D., Reasenberg, P., Jones, L., Raleigh, C. B., Wong, I. G., Scotti, O., & Wentworth, C. (1987). New evidence on the state of stress of the San Andreas Fault System. *Science*, 238(4830), 1105– 1111. https://doi.org/10.1126/science.238.4830.1105